**CSP 595 - Assignment 7**

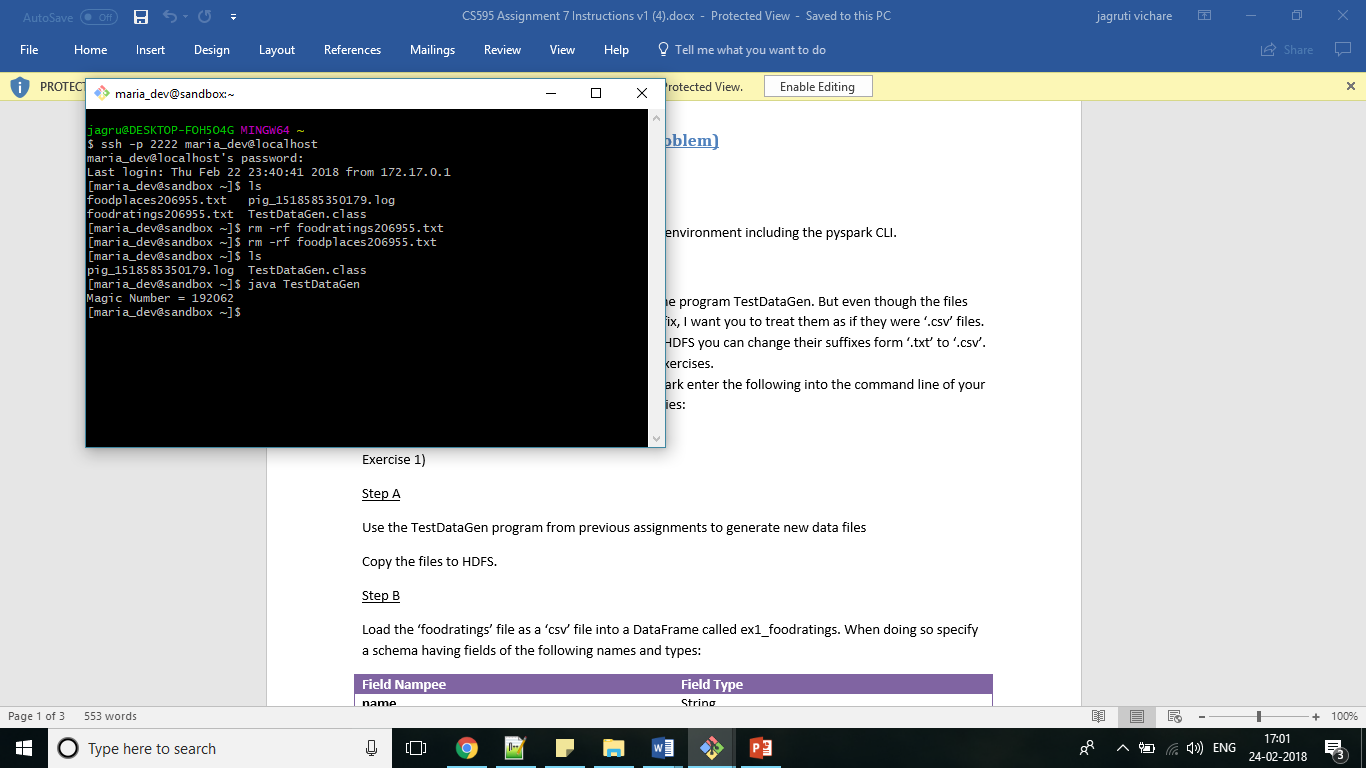
Name: Jagruti Vichare

Email ID: [jvichare@hawk.iit.edu](mailto:jvichare@hawk.iit.edu)

CWID: A20378092

**Exercise 1**

Step A



Step B

from pyspark.sql.types import \*

foodratingstruct = StructType(

[

StructField("name", StringType(), True),

StructField("food1",IntegerType(), True),

StructField("food2",IntegerType(), True),

StructField("food3",IntegerType(), True),

StructField("food4",IntegerType(), True),

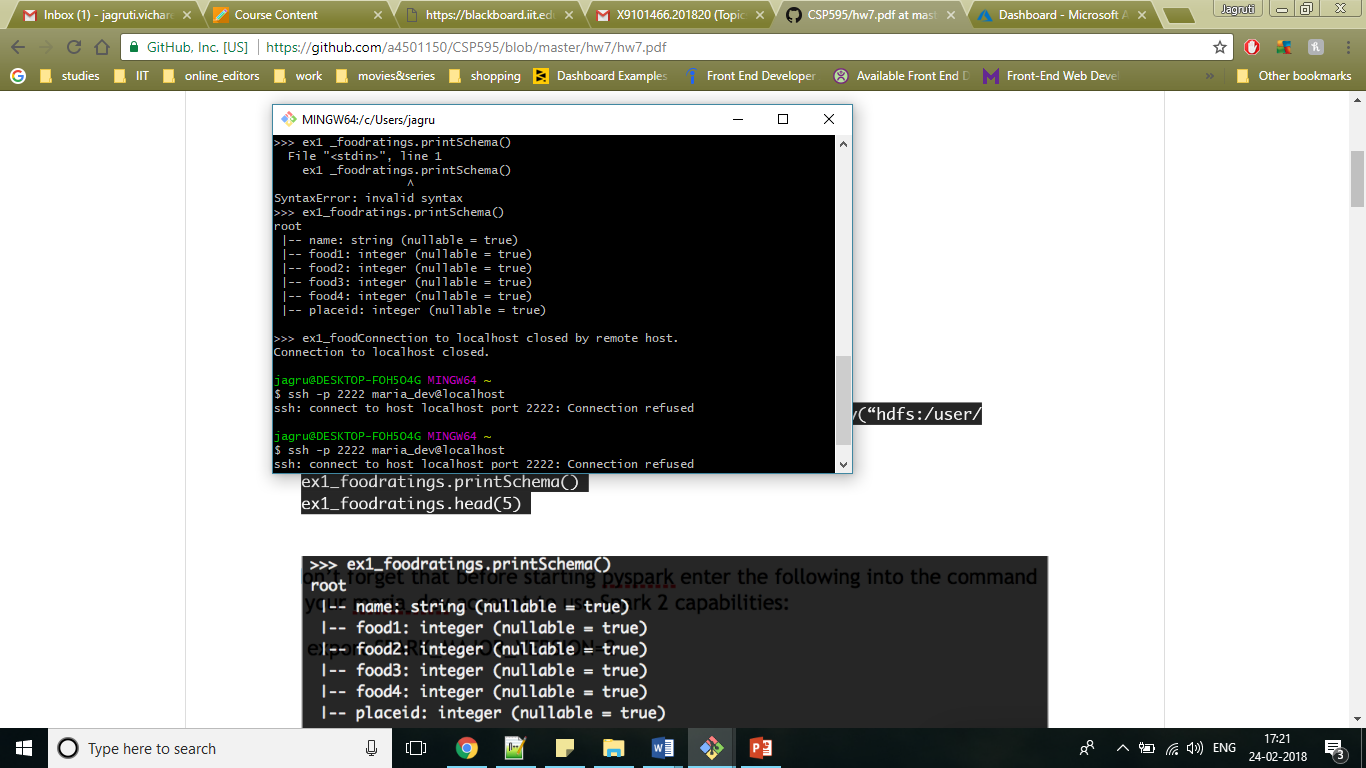
StructField("placeid",IntegerType(), True)

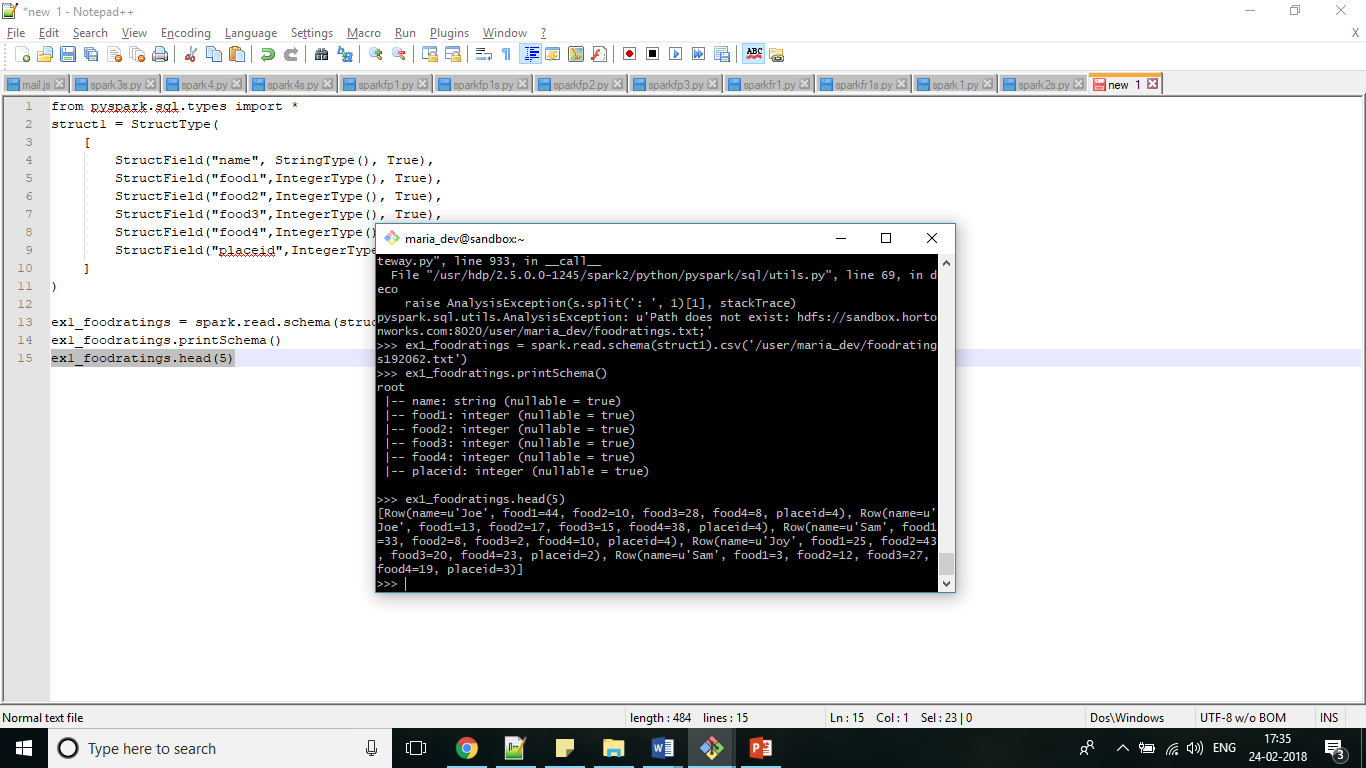
]

)

ex1\_foodratings=spark.read.schema(foodratingstruct).csv('/user/maria\_dev/foodratings192062.txt')

ex1\_foodratings.printSchema()



ex1\_foodratings.head(5)

**Exercise 2**

from pyspark.sql.types import \*

foodplacestruct = StructType(

[

StructField("placeid", IntegerType(), True),

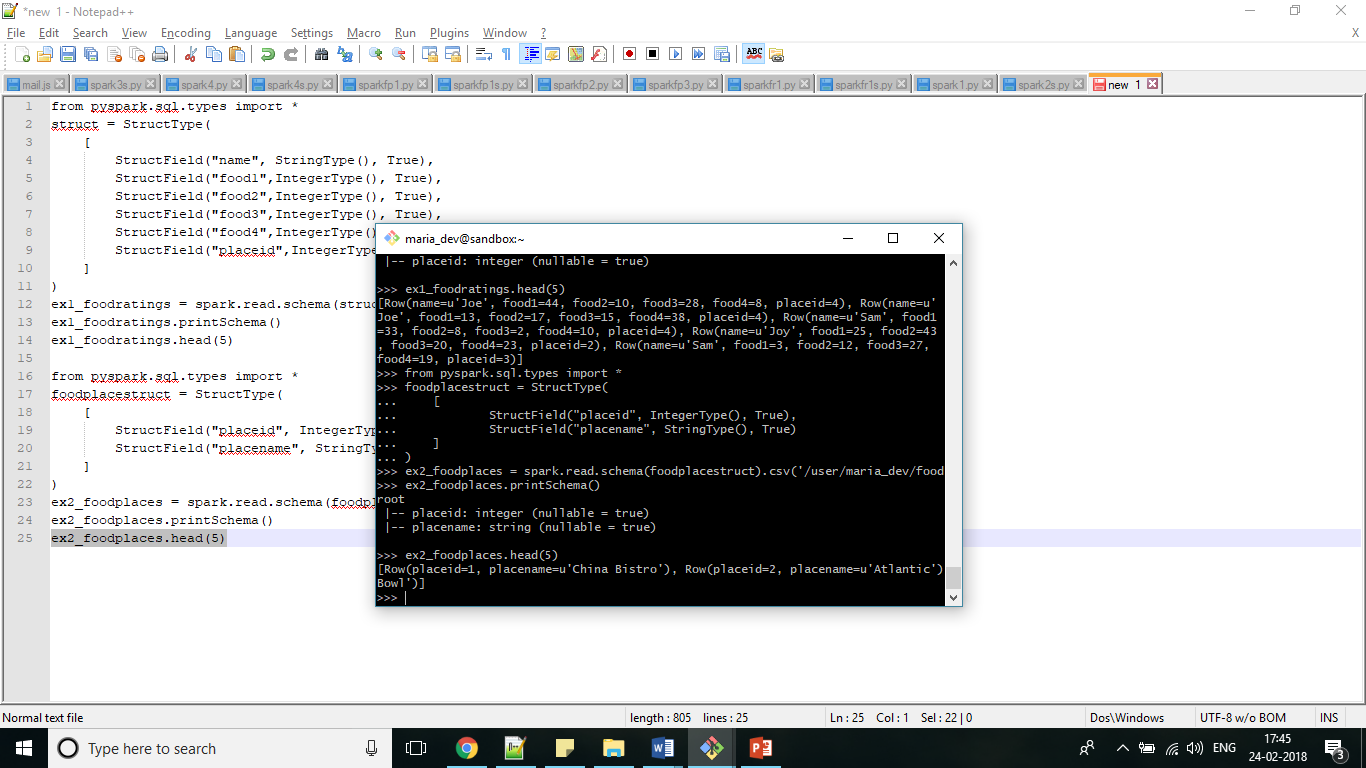
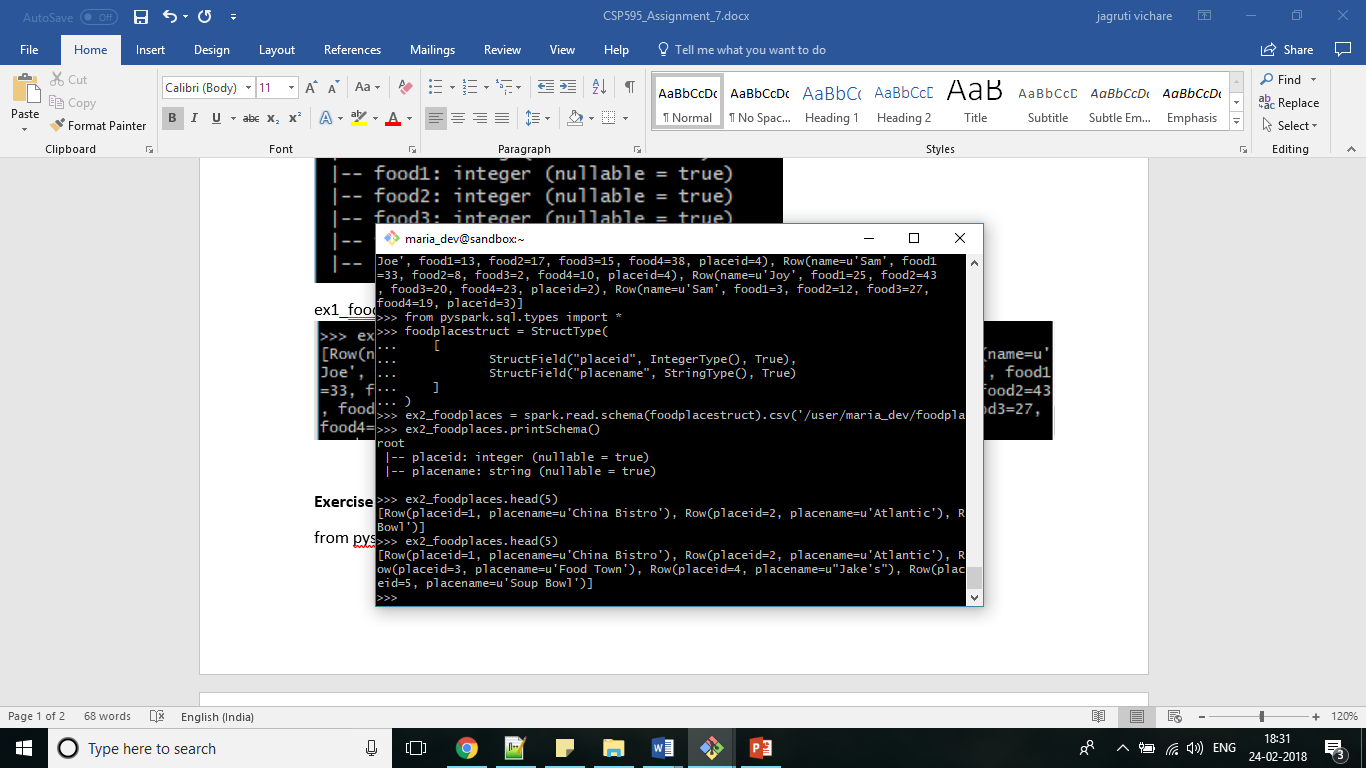
StructField("placename", StringType(), True)

]

)

ex2\_foodplaces = spark.read.schema(foodplacestruct).csv('/user/maria\_dev/foodplaces192062.txt')

ex2\_foodplaces.printSchema()

ex2\_foodplaces.head(5)

**Exercise 3**

Step A

from pyspark.sql.types import \*

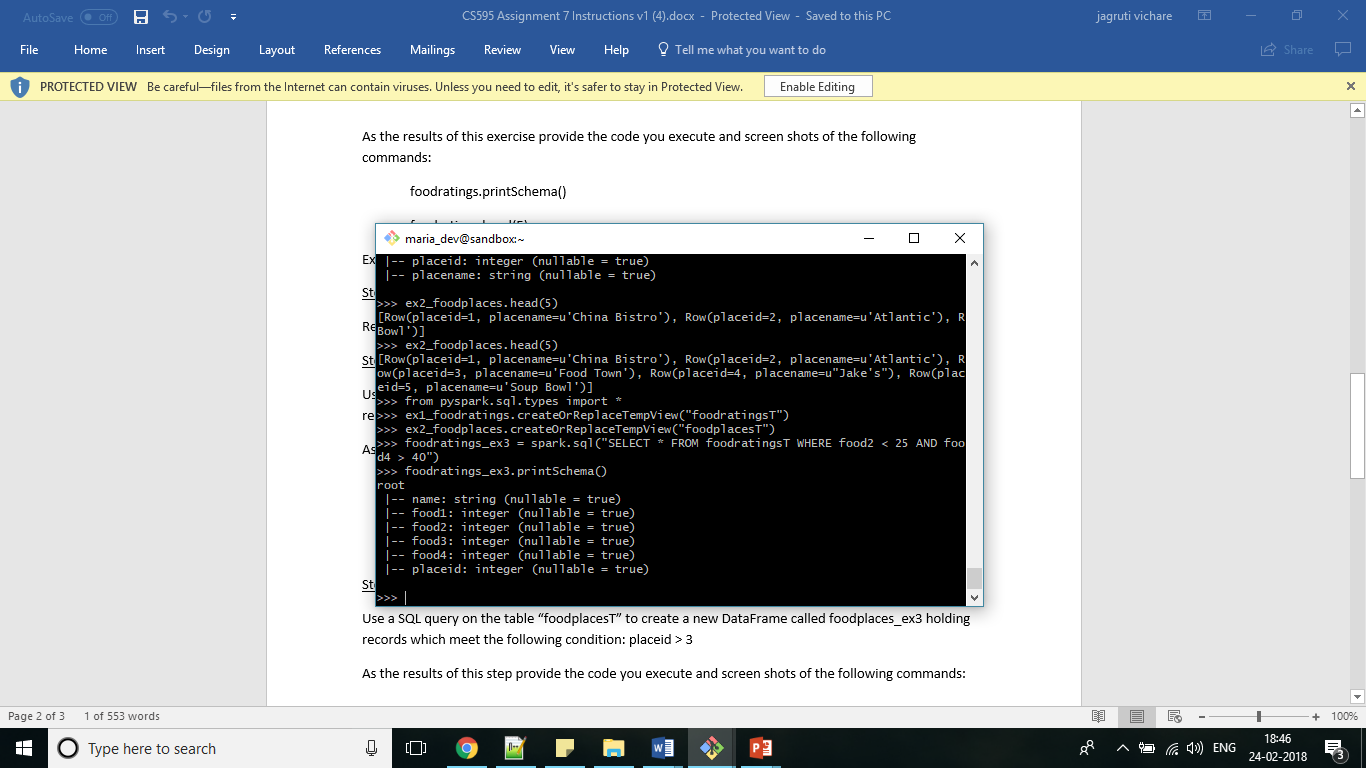
ex1\_foodratings.createOrReplaceTempView("foodratingsT")

ex2\_foodplaces.createOrReplaceTempView("foodplacesT")

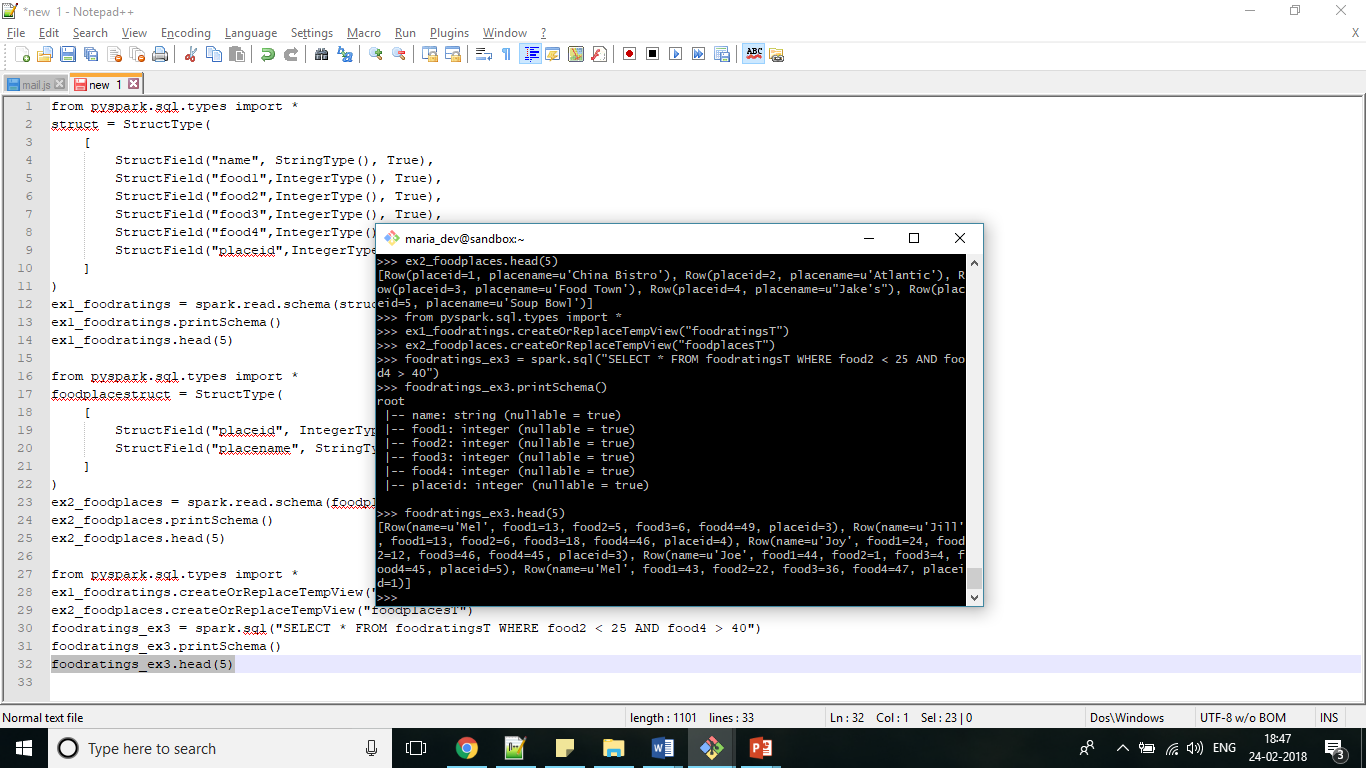
Step B

foodratings\_ex3 = spark.sql("SELECT \* FROM foodratingsT WHERE food2 < 25 AND food4 > 40")

foodratings\_ex3.printSchema()



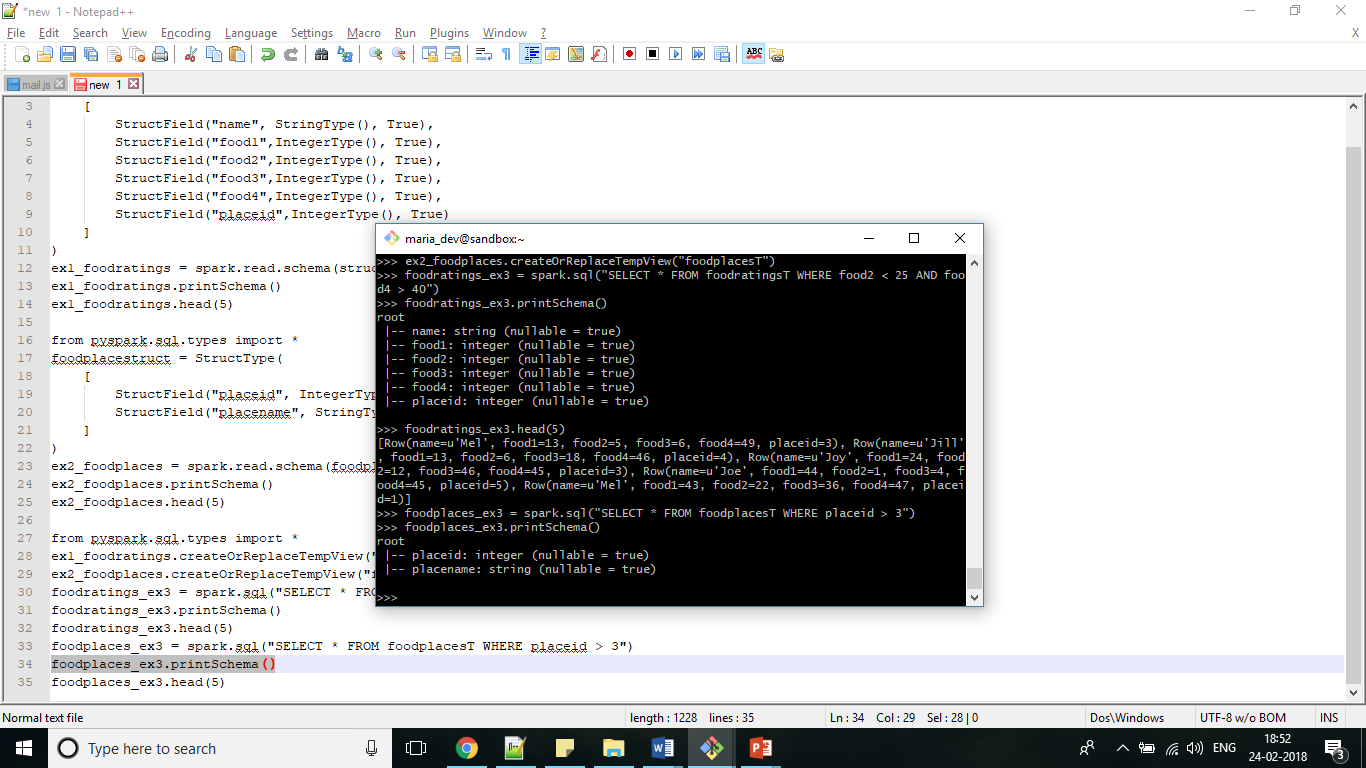
foodratings\_ex3.head(5)



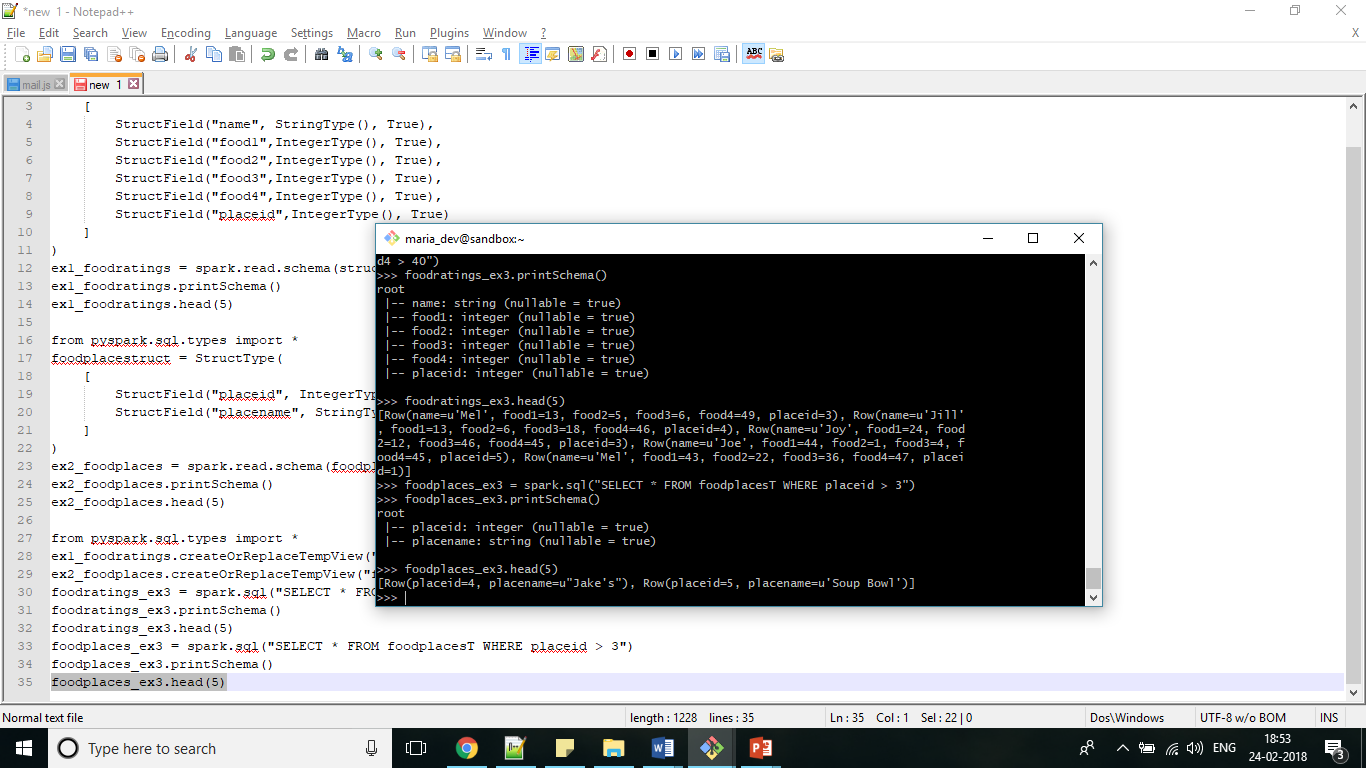
Step C

foodplaces\_ex3 = spark.sql("SELECT \* FROM foodplacesT WHERE placeid > 3")

foodplaces\_ex3.printSchema()



foodplaces\_ex3.head(5)



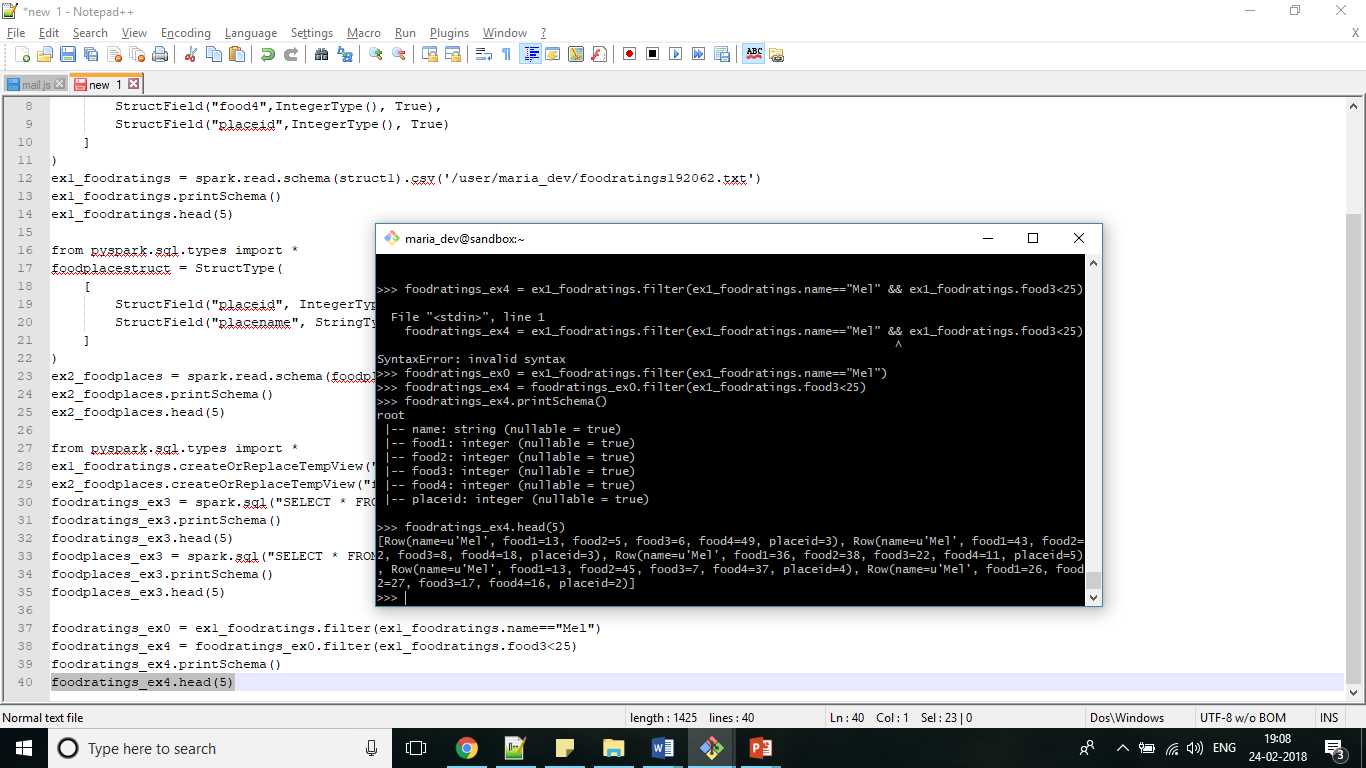
**Exercise 4**

foodratings\_ex0 = ex1\_foodratings.filter(ex1\_foodratings.name=="Mel")

foodratings\_ex4 = foodratings\_ex0.filter(ex1\_foodratings.food3<25)

foodratings\_ex4.printSchema()

foodratings\_ex4.head(5)

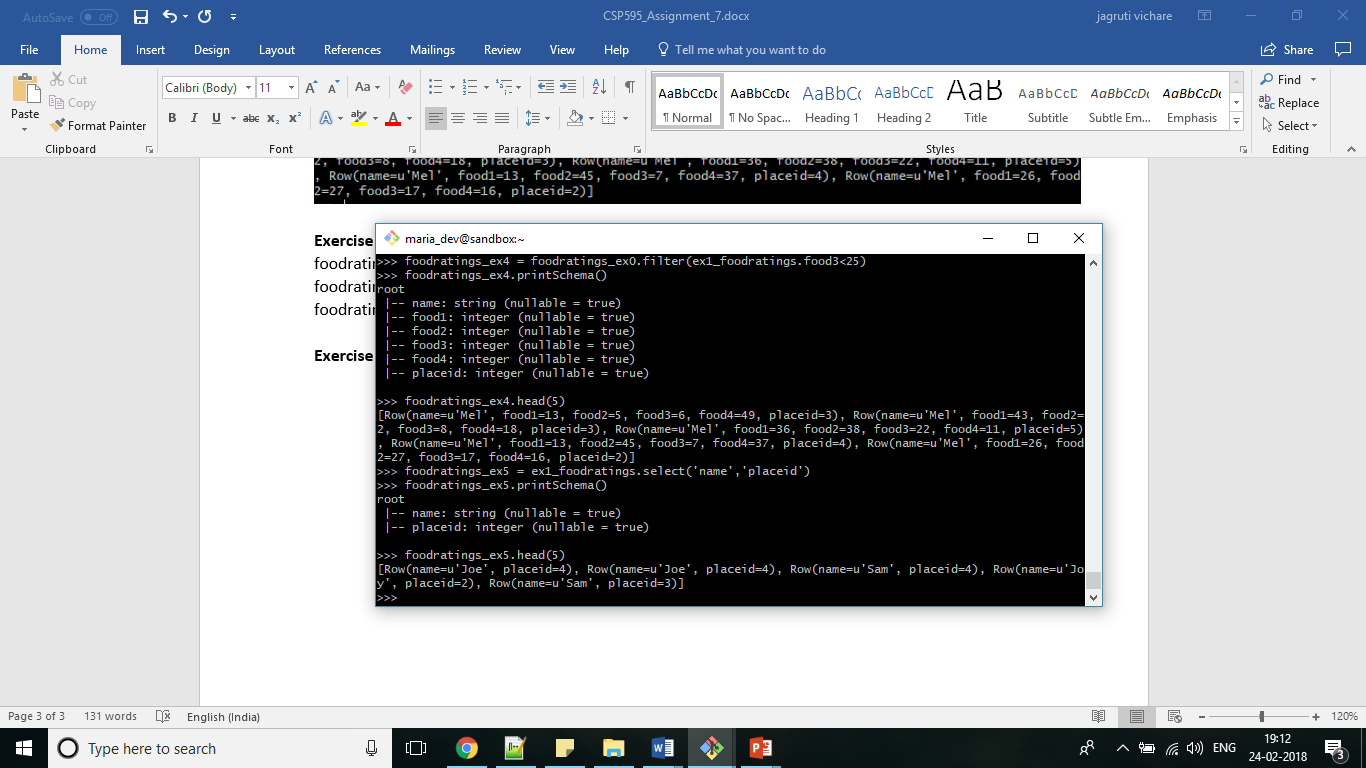


**Exercise 5**

foodratings\_ex5 = ex1\_foodratings.select('name','placeid')

foodratings\_ex5.printSchema()

foodratings\_ex5.head(5)



**Exercise 6**

ex6=ex1\_foodratings.join(ex2\_foodplaces,ex1\_foodratings.placeid==ex2\_foodplaces.placeid,'inner')

ex6.printSchema()

ex6.head(5)

